The Interactional Geometry of a Three-way Conversation

Patrick G.T. Healey (ph@dcs.qmul.ac.uk)

Stuart A. Battersby (stuart@dcs.qmul.ac.uk)
Queen Mary University of London
Interaction, Media & Communication Group
School of Electronic Engineering & Computer Science
London, United Kingdom
London E1 4NS

Abstract

In this paper we describe patterns of spatial co-ordination that, we propose, are a distinctive characteristic of multi-person face-to-face interactions. The data come from a task in which participants describe some simple, non-spatial, computer code in an instructor / learner scenario. Three participants take part; 2 instructors and 1 learner. Using excerpts from these interactions we show that participants make frequent use of combinations of head angle, gesture and participants' positions to 'triangulate' their contributions. We propose that these examples show how face-to-face interaction is distinguished, in part, by the potential it offers for using physical space to create shared 'interactional maps' that provide a structured resource for tracking conversational states.

Keywords: multiparty; face-to-face; head angle; gaze; gesture; interaction

Introduction

Embodied conversation in a shared space is the native habitat for human interaction. It provides people with a rich variety of resources for communication that go well beyond the speech signal including: body position, orientation, gesture, gaze, expression, shared objects, shared spaces and the shared environment.

In this paper we focus on how people deploy gestures and head orientation *in space* to create shared interactional configurations or 'geometries' that specifically exploit the potential face-to-face interaction offers for the simultaneous co-ordination of multiple points in space.

We begin with a brief review of the literature on non-verbal communication and, in particular, a discussion of the kinds of distinctively interactional functions that have been identified for gesture. Experimental evidence shows the importance of these signals for defining, amongst other things, when an interaction starts and finishes, level of mutual-engagement, the roles of each participant (speaker, addressee, over-hearer), their relationship, the shared focus of attention, the boundaries of each turn and the syntactic and semantic organisation of elements within each turn (see e.g. Bavelas and Gerwing (2007); Kendon (1970)).

We argue that although it might appear trivial to observe that face-to-face interaction takes place in a shared space this actually creates an additional set of communicative resources that is quite distinct from visual access *per se*. We go on to present two examples from a corpus of three-person task-oriented interactions that illustrate how people naturally exploit their position with respect to their interlocutors in the shared space to create three-dimensional configurations of gesture and head angle that have distinctive interactional effects.

Non-verbal Interaction

In the literature on non-verbal interaction there is a broad distinction between studies that focus on how individuals produce gestures that are integrated (or not) in various ways with their own speech towards a more detailed consideration of how gestures are deployed in interaction.

Studies of speech-gesture co-ordination often use a 'narrative monologue' paradigm in which participants re-tell a story, typically from a video or cartoon, to camera. This provides a controlled situation in which the relationship between gestures, such as metaphorics or iconics, and the content of the narrative can easily be analysed (e.g. McNeil (1992)). However it also reduces or removes effects of the interaction with an addressee and the dynamics of more open-ended forms of interaction. Particularly clear cases of the kind of phenomena that can be missed by the story telling approach are where participants directly collaborate on the construction and deployment of a gesture (e.g. Furuyama (2000); Tabensky (2001))

Of most interest here are the non-verbal signals which serve, not to convey the topic or content of the discussion, but to manage the interaction itself. This includes body posture (Kendon, 1990), gaze (Bavelas, Coates, & Johnson, 2002; Goodwin, 1979) and gesture (Bavelas, Chovile, Lawrie, & Wade, 1992). Bavelas et al. (1992) coined the term *interactive gesture* to refer to this class of interactive gestures. For example, consider the following extract:

> "and of COURSE there were chances that they would, uh, write something WRONG, you know"?

At the same time as producing 'you know', the speaker points to the recipient with the palm up and all fingers curled except the index finger. Bavelas et. al. suggest that this gesture is equivalent to the verbal 'you know' and means 'do you understand what I am saying'. These gestures occur at a lower rate than those that relate to the topic of discussion and, as Bavelas et al have shown, appear to be exclusive to dialogue.

Goodwin (1979) (see also Bavelas et al. (2002)) described how within a single turn the non-verbal actions of the speaker and the recipient(s) contribute to the incremental co-production of the turn. For example, Goodwin showed that if a recipient is not gazing at the speaker when the speaker gazes at them, this normally leads to a recycling or reconstruction of the utterance. Moreover, if an intended recipient does not return the gaze of the speaker but a third party is gazing at the recipient, the speaker can see the gaze between the third party and the recipient and adjust his own gaze to be towards the third party, upon which mutual gaze between these two participants can be made and successful dialogue continue.

Shared Spaces

An intuitive response to the work cited above is to assume that if people have visual access to each other i.e. if they can see each other's gestures their communication should, all things being equal, be richer, smoother or more effective. However, the extensive literature on video mediated communication demonstrates that the advantage of face-to-face interaction is not due to visual access *per se* (e.g. Whittaker, 2003). There are significant differences between face-to-face interaction and video mediated interaction. For example, participants use more words per turn when communicating over a video mediated communication channel (Whittaker & O'Conaill, 1993), find it harder to spontaneously take the floor requiring more formal handovers (Whittaker & O'Conaill, 1993; Whittaker, 2003) and gestures become exaggerated and mutated from their original forms (Heath & Luff, 1991). These differences are not due to technological factors such as the quality of the video communication channel (Whittaker & O'Conaill, 1993).

The critical point for our argument is that although video connections can provide high-quality, no-delay visual access to an interlocutor they are limited to collections of two-way peer-to-peer channels that do not reproduce the full mutually shared space available in face-to-face interaction. This point is illustrated schematically in Figures 1 and 2. Our claim is that the effectiveness of non-verbal communication techniques depends not only upon visual access, but also on their deployment in a mutually accessible shared space.

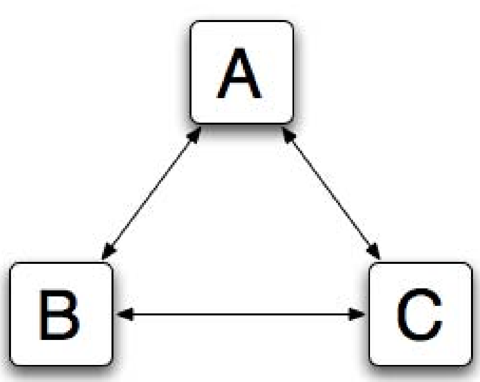

Figure 1: Face to Face Interaction

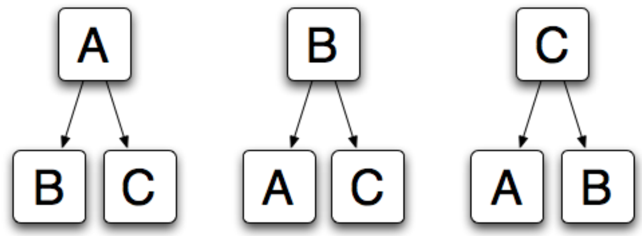

Figure 2: Video Mediated Interaction

Perhaps the most explicitly spatial interactional configurations in the literature are the *f-formation* patterns identified by (Kendon, 1990). This refers to the orientational and postural configurations of participants' bodies during interaction, such that their own individual transactional segments (the area in front of the body in which one carries out their own, individual activities) overlap to form a shared space. It is within this shared space that activities such as gesture can occur. An f-formation shows that we can define an interactional unit spatially, but also it communicates the availability and *exclusivity* of the participant to the interaction. A twisted posture (discussed as 'body torque' by (Schegloff, 1998)) can signal that the participant is still part of the interactional unit, but their attention is temporarily elsewhere (hence loosing the exclusivity of that participant).

Further evidence for the interactional significance of a mutually-shared space is provided by Özyürek who examined the effect of shared space upon gesture formation (Özyürek, 2000, 2002). She was concerned with how addressee location impacted upon the formation of gestures. Her experiments required a participant to recall a scene from a cartoon in which a cat was thrown out of a window to either one or two addressees. In the single addressee condition, the addressee was to the side of the speaker wheres in the two addressee condition they were placed in a V formation (so in front and to the sides of the speaker). It was found that instead of the 'out' gesture being the same for each formation, it was adjusted for each such that the 'outwards' action was out of the shared interactional space. Although this study was not focussed on interactive gestures it shows that participants treat the shared space, or f-formation, defined by their collective body position and orientation as region with special interactional significance.

This point is especially important for multiparty interaction. With the higher number of participants it is spatially more complex, and as the participants are not able to form a standard *viz-a-viz* formation, the interaction makes use more often of areas which are not in the space directly between participants. There is also an added layer of complexity involving mutual knowledge and awareness; two or more people are able to be aware of something in the interaction, whilst there are others who remain unaware. If we refer back to the discussion of Goodwin's work on gaze, we see that what is important is that the speaker can see that the third party is gazing at the inattentive recipient. Over a video channel, this spatial aspect is lost and as such any interactional implications of this are also lost.

We turn now to consider some examples of spatial co-ordination of gesture, head orientation and participant location in a small corpus of 3-way interactions.

Methods

The Augmented Human Interaction (AHI) lab at Queen Mary houses an optical motion capture system. The system consists of an array of infra-red cameras and, by attaching reflective markers to the bodies of participants, we are able to track their movements in three dimensional space through the motion capture system. This technology allows us to analyse the participants' movements in more detail than traditional video alone as we have the precise 3D coordinates of discrete segments of the body which preserves the spatial nature of the interaction. This three dimensional data can be reconstructed as wireframe representations of the body and analysed along with video and audio data (see Battersby, Lavelle, Healey, and McCabe (2008) for more detail). In the current study these have been used to aid in the analysis, however the data is displayed as images from traditional video. Three camera angles were used, one above and one either side of the participants.

Task Description: The data reported here is is drawn from a corpus of exploratory studies of multi-party interaction which have taken place in the AHI lab. In these studies three participants take part in three rounds of collaborative interaction. On each round two participants, the 'instructors', are given one printed copy of a Java application with its associated class hierarchy. Each class is printed on a separate sheet of paper. Before the round begins the instructors are asked to discuss the code together, make sure they both understand it and then return the printed code to the experimenter. The third participant, the 'learner' then joins them and they sit on pre-positioned stools in a circle (see e.g. Figure 3).

The instructors then have as long at they like to explain the code to the learner, notifying the experimenter when they are satisfied that the tuition is complete. No restrictions are placed on how they explain the code, or on the interaction other than they must not use pen and paper. To provide participants with a criterion of understanding and to assess the learners comprehension at the end of each round the learner was asked to reassemble the hierarchy on their own using a drag and drop interface on a laptop.

Three different Java class hierarchies were created: 'Student', 'MP3 Player' and 'Retailer'. These materials were designed to embody hierarchical relationships but without involving any simple spatial relations (such as those involved in describing a route or the layout of a kitchen). A different set of Java materials is used on each trial, with order of presentation controlled across groups of participants, and roles are systematically changed so that each participant acts as the learner once.

Participants: To ensure familiarity with Java all the participants were recruited from amongst Computer Science undergraduate or postgraduate students at Queen Mary. Three groups of three participants produced the data reported here. All groups were arranged by the experimenter and consisted of 7 male and 2 female participants, aged between 20 and 30.

Using Space in Interaction

Overall the three groups of participants produced a total corpus of 25mins 15secs of video and 3D motion capture recordings with the average length of each round of 2 mins 48secs.

Our primary focus here is on patterns of interaction in this sample that are constitutively spatial in character i.e., that make direct use of the potential face-to-face interaction offers for co-ordinating multiple communicative resources in space. Before moving on to this we first provide a brief overview of the way participants approached the task and their general use of gesture and body orientation.

In a typical round one instructor would take the lead in describing the code to the learner, checking with the other instructor at various points and occasionally directly inviting comments or elaboration from them. The 'second' instructor would sometimes interject during these expositions or, more often, provide additional clarification or elaboration at the end of the exchange. Both instructors' head orientation and contributions during these explanations were predominantly oriented towards the learner, even when not taking a primary role in the explanation. However, during checks or interjections the instructors would normally briefly orient to each other before returning their attention to the learner. The learners typically adopted a relatively passive role. Sometimes asking for clarification during the initial explanation but more often waiting until the instructors signaled that their description of the code was complete. Learners would then read back there understanding of what had been presented with both instructors providing feedback.

Simple Uses of Gesture and Orientation

During these interactions the speaker (instructor or learner) produced most of the gestures and the silent participants normally kept their hands resting on their knees or in their lap (see e.g. Figure 4 and Figure 5).

A common feature of gesture in these interactions was instructors would use different spatial locations between the participants to represent different elements of the hierarchy. These iconic gestures were not only restricted to a personal gesture space in front of the speaker, but in some cases also extended into the shared interaction space.

Figure 3 illustrates an example of the learner sharing an iconic gesture space directly with an instructor. One instructor has his back to the camera, with his fellow instructor to his left and the learner in front of him. This image shows the instructor drawing out the hierarchy, and the learner co-referencing this hierarchy (see Furuyama (2000) for similar examples from instructional dialogues). Notice that here the participants are using the shared gesture as the spatial frame of reference and not, for example, their own body centered co-ordinates which would involve a left-right reversal of the hierarchy.

Figure 3: Shared Gesture Space

In addition to these primarily iconic gestures, participants also used familiar non-verbal interactional cues. For example, using their gaze (both with and without accompanying speech) to address people, seek confirmation and seek clarification and using interactive gestures such as short hand flicks to reference another person while speaking (see e.g. Bavelas and Gerwing (2007); Kendon (1970)).

Triangulated Uses of Gesture and Orientation

The discussion so far has highlighted the use of gesture and/or orientation in a shared space and the parallels with previous work on non-verbal interactions described in the introduction. We turn now to patterns of non-verbal interaction that make essential use of the spatial arrangement of participants. In the context of this task these patterns most commonly occurred where participants engaged a third party, typically the other instructor, while addressing a second party, typically the learner. We gloss these patterns of simultaneous engagement as moments of *triangulation* to highlight the way that they make direct use of the mutually-known spatial arrangement of the three participants *in space*.

Restricting our attention to head and gesture orientation there are three basic spatial possibilities: 1) the speaker orients to the third-party with a gesture while continuing to orient to the addressee with their head; 2) the speaker orients to the addressee with a gesture and orients to the third party with their head or 3) the speaker uses a combination of head and gesture orientation to the third party.

An example of the first of these three possibilities is given in Figure 4. Here Instructor 1 is in the middle of the scene facing the camera. On the left of the image is Instructor 2, and on the right of the image is the learner. Prior to this situation, Instructor 2 had described the class hierarchy to the learner. Instructor 1 then takes over and explains how the application works and makes use of the classes described by Instructor 2 saying "l-lower hierarchy classes like the masters and undergraduate". At the start of this turn he is facing the learner and not gesturing but as he reaches "masters" he creates a left handed palm up gesture towards Instructor 1 while maintaining shared gaze with the learner.

Figure 4: Divergent Orientation of head angle and Gesture

A more complex example that illustrates the third possibility is provided by the sequence illustrated in Figures 5, 6 and 7. Here, Instructor 1 (female, white cap) is explaining some of sections of code (methods) which are part of the Playlist class to the learner (male, black cap). As she does this she gestures with both hands; the left hand is held out between her and the learner with her fingers extended, the right hand is counting along the fingers (by pointing at them) as she lists each method. She is also looking (gazing) towards the learner. The learners hands are resting on his legs and he is back channeling verbally and with head nods. The second instructor (male, blue cap) is looking towards the learner but not gesturing or speaking.

Figure 5: Listing Methods to the Learner

She continues her list with "add to track" but then initiates a repair on this to change it to "add track". In the middle of this repair she poses the question "is it add track". As she says this her gesture & head configuration changes; her head turns towards Instructor 2, and her right hand moves from being a counter on the left hand to a point in the direction of Instructor 2. Instructor 2 responds by changing his head orientation towards Instructor 1. Note, however, Instructor 1

Table 1: Excerpt from Example 2 “Add track”

Instructor 1: “add play playlist” (0.1) (Fig. 5)
Head: Towards the learner, angled down slightly
Gesture: Left hand is between herself and the learner with fingers extended. Right hand is counting along the left hand’s fingers with a pointing gesture
Instructor 1 : “add to track” (0.1)
Instructor 1: “is$_1$ it add track”$_2$ (0.3) (Fig. 6)
Head$_{1\&2}$: Turns from learner to face Instructor 2
Gesture$_2$: Left hand stays stationary between herself and learner, right hand moves to be placed between herself and Instructor 2, pointing towards Instructor 2
Instructor 1: “I think$_1$ add track”$_2$ (Fig. 7)
Head$_1$: Turns from Instructor 1 to face the learner
Gesture$_1$: Right handed point turns in an arced motion around her body to be placed between herself and the learner, now pointing at the learner. Left hand remains in place.
Gesture$_2$: Left handed gesture ends and rests on the leg.

continues to hold the floor, actually answering the question herself as she turns her head back to the Learner (Figure 7).

During this sequence her left ‘list’ hand maintains it’s position and shape between herself and the learner. Thus, although she has temporarily changed orientation to her co-instructor, it appears natural to interpret her left hand as helping to maintain both the relevance of the business of listing the methods while the repair is completed and to maintain the fact that this list, unlike the repair that is part of it, is primarily addressed to the Learner. The effect is one of keeping the conversation between herself and the Learner ‘on hold’ and limiting the rights of the co-instructor to take the floor.

Figure 6: Repairing “Add to track”

Instructor 1 answers her own question with “I think add track” as she completes “think” she synchronously turns her gaze and point (in a slightly arced trajectory around herself) towards the Learner. As she is doing this, i.e. after she has turned away, the co-instructor confirms her repair with “yeah add track yeah”. As her right handed point comes into the space between herself and the learner her left hand drops from its holding position, the learner turns his gaze back towards her and the instructional dialogue continues. It is worth emphasizing the fact that it is only after Instructor 1 turns her right handed point back towards the learner that left handed ’hold’ gesture is dropped.

Figure 7: Resume

Our claim is that this example illustrates how participants can make simultaneous use of two different spatial orientations, in this case involving two gestures with different forms and orientations to the co-instructor and to the Learner, as a resource for managing the interaction. In particular, as a resource for understanding the incremental structure of the turn under construction and the varying status of each participant with respect to those increments (cf. (Goodwin, 1979))

The Distribution of Triangulations

In order to get a better understanding of the distribution of these patterns of triangulation the video recordings were transcribed and coded using Elan. All occurrences where participants created one of these configurations of gaze and gesture were coded. The coding focussed only on situations where there was a visible change in the orientation of the head and/or hands of the speaker relative to the other participants *during* the production of a turn and not at turn completion points or during explicit hand-over of the floor. In these cases we coded whether the gesture and/or head orientation changed or maintained. We did not include cases if only the head moved whilst the hands were ‘resting’, either placed on the knees or clasped together in the lap and not gesturing

This coding yielded 61 cases of triangulation of which 44 were where the hands and head had divergent orientations and 17 where they were combined. This indicates that this use of shared space, at least in the kinds of interaction we studied is relatively common with one instance of an observable triangulation occurring approximately every 25 seconds of interaction.

It is also clear from this data that there is a systematic contrast in the significance of these movements for the primary addressee and the third party. In 63% of cases the third party responds to the change of orientation by turning away from the primary addressee and towards the speaker. By contrast the primary addressee changes orientation away from the speaker (and towards the third party) in only 16% of cases.

Conclusion

The central claim of this paper is that people use their physical arrangement in space during multi-party interactions as an additional specialised resource for communication. More specifically, that the mutually-known arrangement of participants, gestures and orientation in physical space can be used to create and manage what amount to 'interactional maps' of how each contribution is constructed, maps of the relationships between different contributions (and elements of contributions) and maps of different people's stance toward or participant role with respect to those contributions.

As the contrast with video-mediated communication highlights, this is more than a matter of being able to see the other participants. It critically depends on the fact that the participants are in the same shared space and that they mutually know this is the case.

The evidence from the present study suggests that when people have the opportunity they naturally make extensive use of this potential. We can only speculate on the function of these different interactional geometries but our study minimally suggests that they support, amongst other things, references to locations as representative of prior turns and as representative of the producers of those turns. They also support patterns of turn management, helping participants to keep sub-dialogues visually (as well as structurally and semantically) distinct.

We have described these configurations as constitutively spatial. What we mean by this is that they involve a use of space that could not, in principle, be reproduced, for mutliparty interactions using point-to-point video. This may help to explain the paradoxical finding in the literature that, while face-to-face is clearly better than audio only interaction, there is little or no difference between video and audio only. It appears that it is not the visual channel *per se* but the shared space itself which provides the critical interactional advantage.

References

Battersby, S. A., Lavelle, M., Healey, P. G., & McCabe, R. (2008, May). Analysing interaction: A comparison of 2d and 3d techniques. In *Conference on multimodal corpora.* Marrakech.

Bavelas, J. B., Chovile, N., Lawrie, D. A., & Wade, A. (1992). Interactive gestures. *Discourse Processes*, *15*, 469-489.

Bavelas, J. B., Coates, L., & Johnson, T. (2002). Listener responses as a collaborative process: The role of gaze. *Journal Of Communication*, *52*, 566-580.

Bavelas, J. B., & Gerwing, J. (2007). Conversational hand gestures and facial displays in face-to-face dialogue. In K. Fiedler (Ed.), *Social communication.* Psychology Press.

Furuyama, N. (2000). Gestural interaction between the instructor and learner in *origami* instruction. In D. McNeil (Ed.), *Language and gesture* (p. 99-117). Cambridge University Press.

Goodwin, C. (1979). The interactive construction of a sentence in natural conversation. In G. Psathas (Ed.), *Everyday language: Studies in ethnomethodology* (p. 97-121). Irvington Publishers.

Heath, C., & Luff, P. (1991). Disembodied conduct: Communication through video in a multi-media office environment. *Conference on Human Factors in Computing Systems, Proceedings of the SIGCHI conference on Human factors in computing systems: Reaching through technology*, 99-103.

Kendon, A. (1970). Movement coordination in social interaction: Some examples described. *Acta Psychologica*, *32*, 100-125.

Kendon, A. (1990). *Conducting interaction: patterns of behavior in focused encounters*. University of Cambridge.

McNeil, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago Press.

Özyürek, A. (2000). The influence of addressee location on spatial language and representational gestures of direction. In D. McNeil (Ed.), *Language and gesture* (p. 64-83). Cambridge University Press.

Özyürek, A. (2002). Do speakers design ther cospeech gestures for their addressees? the effects of addressee location on represetational gestures. *Journal of Memory and Language*, *46*, 688-704.

Schegloff, E. A. (1998). Body torque. *Social Research*, *65*, 535-596.

Tabensky, A. (2001). Gesture and speech rephrasings in conversation. *Gesture*, *1:2*, 213-235.

Whittaker, S. (2003). Theories and methods in mediated communication. In A. C. Graesser, M. A. Gernsbacher, & S. R. Goldman (Eds.), *Handbook of discourse processes* (p. 243-286). Lawrence Erlbaum Associates Inc.

Whittaker, S., & O'Conaill, B. (1993). An evaluation of video mediated communication. In *Chi '93: Interact '93 and chi '93 conference companion on human factors in computing systems* (pp. 73–74). New York, NY, USA.